From Tool Use to Social Interactions

Anna Strasser

1 Introduction

Human-machine interactions are often exclusively described as tool use and not as social interactions. However, this contradicts experiences we can make in interactions with artificial agents, especially when social robotics come into play. Without a doubt, humans have the ability to bond emotionally not only with living beings but also with inanimate artificial agents. Consequently, we treat artificial agents as if they were social agents (cf. Carpenter 2016; Darling 2016). However, behaving socially in front of artificial agents might not yet fully justify qualifying them as social agents and attributing social agency to them.

This tension provides a motivation to question the widespread restrictive conception of sociality, in particular of social agency, which excludes inanimate objects as potential participants in social interactions. According to this view, it seems widely accepted that being alive constitutes a minimal condition (necessary but not sufficient) to be considered a social agent. Following this line of thinking, it is argued that artificial agents cannot qualify as participants in social interactions because they lack the essential properties of living beings. Particularly in the debates about joint actions in the field of philosophy of mind, it becomes obvious that definitions of socio-cognitive abilities that lay a foundation for the attribution of social agency exclude artificial agents from the outset.[1]

1 I am not questioning that humans can behave socially toward all sorts of entities. However, I think that this should be separated from the question of whether it is justified to attribute socio-cognitive abilities to these entities. At this point, it should be noted that in other fields of research the notion of 'social interaction' seems to be understood more broadly, namely in the sense that everything is understood as social interaction that follows a specific pattern of human behavior. Thus, all interactions in which we

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - CC BY-SA

In fact, one can observe that notions describing agency, the ability to act jointly, or mindreading are restrictive even in a more radical sense (cf. Davidson 1980; Bratman 2014; Fodor 1992). They characterize such abilities as if they were unique to sophisticated human beings only. One factor contributing to restrictive conceptions might be that philosophy's main objective is to create sharp, clear-cut notions, resulting in a general tendency to focus on ideal cases with rather demanding conditions.

Therefore, it is not surprising that some sophisticated terminology of philosophy already reaches its limits when it comes to socio-cognitive abilities of other living agents, such as children or non-human animals (cf. Brownell 2011; Heyes 2014, 2015; Pacherie 2013; Perler 2005; Premack/Woodruff 1978; Vesper et al. 2010; Warneken et al. 2006). In contrast, developmental psychology and animal cognition demonstrate gradual appearances and multiple realizations of socio-cognitive abilities, but these cannot be captured by the aforementioned sophisticated terminology.

Proposals responding to these shortcomings introduce terms such as simple forms, proto-cases, or quasi-states to extend the restrictive conceptual framework to a wider range of cases. For example, Perler and Wild talk of non-human animals as having simple thoughts – "simple 'Geisthaber'" (2005: 70) – to describe the thinking abilities of non-human animals.

So-called minimal approaches follow a comparable strategy, and I aim to show that their strategy offers a promising starting point for characterizing the abilities of artificial agents that cannot be captured by the sophisticated terminology. By challenging overly demanding conditions of certain philosophical notions, minimal approaches can cover a broader range of socio-cognitive abilities. Examples are notions such as *minimal mindreading*, *minimal sense of commitment*, and *shared intentions lite* (cf. Butterfill/Apperly 2013; Michael et al. 2016; Pacherie 2013).

Also in the debates of moral philosophy, there are more and more positions that question all too demanding conditions regarding (moral) agency. Especially with respect to the evaluation of interactions with potentially social robots, arguments are made for an extension of the conceptual framework (Wallach/Allen 2009). This becomes particularly evident concerning concepts that describe assumed unique human capabilities such as autonomy. Thus,

behave socially towards entities count as social interactions. This also means that artificial systems can trigger social behavior, regardless of whether a social agency can be attributed to them.

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

in debates about moral agency in general, a more fine-grained differentiation in terms of autonomy is proposed (Darwall 2006: 265). Instead of necessarily presupposing full-fledged autonomy, several degrees of autonomy are conceptualized. Furthermore, the necessity of other demanding conditions is questioned. For example, Floridi and Sanders (2004) claim that consciousness, intentionality, mental states, and intelligence are not necessary for characterizing moral agency of artificial agents. According to them, it is sufficient to require negative autonomy (absence of external force and control), interactivity, and adaptability (see also, Wallach/Allen 2009; Misselhorn 2018).

Reflecting on the increasingly important role artificial agents play in our social life, controversial positions have been developed when assessing the status of artificial agents. Aside from an instrumentalist view, which is based on a demanding understanding of agency and claims that artificial agents can, in principle, only be considered as tools (cf. Johnson 2006), one can now find positions arguing that artificial agents can be considered as social agents (cf. Darling 2016; Hakli/Seibt 2017). Assuming that the impact of human-machine interactions on our social life will increase and that such interactions, to most appearances, will significantly differ from other tool use cases, this paper aims to elaborate under which circumstances we are justified to consider artificial agents as social agents instead of mere tools.

Arguing for the claim that one should broaden the conceptual framework of social agency if one is to arrive at adequate characterizations of all varieties of social interactions, I use the debate about joint actions in philosophy of mind as an example to explore the extent to which one can ascribe a minimal form of agency to artificial agents that do not meet the demanding conditions of full-fledged agency. The suggested conceptual framework allows for fine-grained distinctions between mere behavior and action (Strasser 2006: 172). Since agency alone is not sufficient to qualify as a social agent in joint actions, I also investigate how to conceptualize the necessary socio-cognitive abilities of artificial agents in order to ascribe social agency to them.

The decision to deal with joint actions should not be understood as an equation of joint actions and social interactions. Joint actions are only one form of many forms of social interactions. They are perhaps one of the most demanding forms of social interaction, and this seems to me to be a particular challenge. My concern is to show precisely for sophisticated social interactions to what extent one can argue that artificial agents can also be qualified as possible participants here. Moreover, the debate about joint actions provides a template for rethinking anthropocentric assumptions. In this context,

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Christian List (2021) points out an interesting parallel regarding group action and artificial agents. According to him, both phenomena can be understood as interactions with non-human, goal-directed agents (collective agency/artificial agency) that can change the social world. And this can be taken as a reason to reconsider some of our anthropocentric moral assumptions.

Although both conceptual issues and psychological factors contribute to a theoretical foundation on which we can argue for the potential sociality of artificial agents, this paper focuses on the conceptual issues. That is, I am less concerned with psychological factors that contribute to the subjective impression that certain interactions with artificial agents are social interactions than I am with elaborating minimal necessary conditions that must be met by artificial systems to be considered proper participants in social interactions. Therefore, I am primarily addressing the conditions that artificial agents must satisfy so that we are justified in ascribing social agency to them. This may help provide a fully comprehensive basis for an analysis of all factors that contribute to transforming human-machine interactions from the mere use of tools into social interactions.

Questioning that all human-machine interactions can be reduced to mere tool use, I point to social phenomena in which artificial agents are not merely involved but play an active role and show that established notions from the philosophy of mind do not provide an adequate conceptual framework for such phenomena (section 2). To conceptualize these phenomena, one must rethink restrictive conceptions of sociality. Thereby, questioning whether biological constraints inhibit a necessity will be of primary importance (section 3). Understanding joint actions as one interesting and challenging case of a social interaction, I suggest a conceptualization of a non-human-centered version of acting jointly that will specify the conditions artificial agents must meet in order to enter the space of social interaction (section 4). Finally, I point to research findings that indicate to what extent proposed conditions are already met by actual artificial agents (section 5).

2 Varieties of human-machine interactions – the discovery of terra incognita

Humans have entertained manifold types of tool use from the Stone Age to the present day, and tools have changed over time. In general, tool use is understood as the instrumental use of objects to achieve certain goals. For ex-

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

ample, using a hammer to drive in a nail is a prototypical case. However, not only inanimate objects can be used as tools, but also humans or other living beings can be used as tools. Being a social agent does not prevent you from being used as a tool. Animate and non-animate entities can be used as tools. However, to be considered an active participant in a social interaction, one must qualify as a social agent, whereby I do not presuppose that all social agents must be animate entities. In our time, there are increased actions involving various types of new technological devices, such as the Internet, cell phones, social networks, and last but not least, chatbots and social robots. In fact, artificial intelligence is shaping many of these new technological devices.

The question I am concerned with is whether all these devices should be considered mere tools. Obviously, tools differ concerning their complexity; we can easily distinguish simple tools such as a hammer from more complex tools such as statistical software or social robots. However, in view of technical devices that display some learning abilities, the differences become greater. Normally, tool use implies to some extent that the user is in control of the tool. However, there are tools that are only to a very small extent under our control; such tools show some degree of autonomy or are even able to adapt and learn. This is reflected, for example, in the distinction between so-called *in-the-loop systems*, *on-the-loop systems*, and *out-of-the-loop systems* (cf. Leveringhaus 2016: 3; Loh 2019: 33). The former are subjects to human control, whereas *out-of-the-loop systems* describe machines in which humans even do not have an intervention option. In-between, there are *on-the-loop systems*, which have some autonomy, but the human still can intervene. On-the-loop systems can decide and act autonomously, but the final decision remains with the human and thus, it is argued, the responsibility. With respect to the different grades of autonomy, some tools already have agent-like properties. This means that there are interesting differences on the table that require finer conceptual differentiation and may raise the question of whether some tools can not only qualify as agents but also as social agents.

The reasons why describing human-machine interactions as mere tool use contradicts the experiences we make with artificial agents are manifold. Apart from the hypothesis that artificial agents may indeed qualify as a new type of social agents, which is this paper's focus, psychological factors also contribute to the impression that certain human-machine interactions are social interactions. For example, interactions with social robots are perceived as social and not as tool use because these products are specifically designed to trigger the human tendency to anthropomorphize (in the sense of a tendency to-

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - CC BY-SA

wards behaving socially). Treating entities socially, we behave as if they were social agents. This can be illustrated by our relation to certain artificial toy pets. Darling (2016) reports from a group of participants that were given a cute robotic dinosaur called *Pleo* to interact with. In the end, they were asked to tie up, strike, and *kill* their *Pleo*. However, due to the human tendency to sociality – in that case, an animomorphization – participants refused to *hurt* their robot. The human ability to emotionally connect not only with living creatures but also with inanimate artificial agents plays an important role in many domains, even in the military sphere, as Carpenter (2016) reported. Undoubtedly, we often interact with artificial agents as if they were social agents and not just tools.

At this point, some positions argue that artificial agents (or actually their creators) simply trick humans into attributing properties to artificial agents falsely and thus conclude that all human-machine interactions are, in the end, just tool use, no matter how social such interactions appear (Bryson 2010). To this end, it might be mentioned that anthropomorphism, including animomorphism, is traditionally seen as a bias, a category mistake in psychology (cf. Damiano/Dumouchel 2018; Loh 2019). However, admitting that there are cases of being tricked does not exclude the possibility of genuine social interactions with artificial agents. Given that some human-machine interactions, especially those with social robots, are significantly different from those that constitute mere tool use (e.g., by involving only simple tools, such as laptops or toasters), I argue that it is reasonable to address these differences by re-examining the conditions artificial agents must fulfill in order to display sociocognitive abilities and, thus, be considered as a new type of social agent.

Of course, observing a person who spends most of her time with a care robot, one can describe many of the involved interactions as tool use. The person uses the robot as an assistant (tool) that gives support and takes care of activities the person is not able to do. Nevertheless, it is at least conceivable that the person may also communicate with such a robot and satisfy social needs. Regarding those interactions, the question arises whether the care robot takes on the role of a social agent here. It is not necessary to go so far and claim that a care robot could replace a human caregiver. However, it seems reasonable to assume that some of those interactions have the potential to lead to experiences which are strikingly similar to those we make with human counterparts, and this might be due to abilities the robot actually has.

These considerations suggest that not all interactions with social robots should be reduced to instrumental, ordinary tool use. When talking about

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - CC BY-SA

tool use, one usually assumes that tools are rather passive. That is, we do not assume that our tools are capable of acting autonomously, nor do we expect them to adapt to our behavior or even learn new behaviors and respond to social cues. At least in the last decades, we would have been quite irritated if hammers or bicycles would suddenly smile or say something back. Whereas social bots are explicitly developed to be companions that adapt, learn, and communicate. Some of them are able to process and display social cues.

Moreover, the assumption that certain human-machine interactions are in some ways similar to human-human interactions has already found its way into empirical research. Here, it is assumed that the way humans behave in interactions with artificial agents bears at least some resemblance to the way they behave in interactions between humans because people make socialness attributions (cf. Hortensius et al. 2018). Consequently, experimental protocols with artificial agents are used to gain insightful information about humans' social cognitive mechanisms (cf. Wykowska et al. 2016). If there were no similarities at all, such experiments could not provide any information about human-human interactions.

Although psychological factors help characterize specific features of human-machine interactions, it is important not to rely exclusively on first-person attributions when arguing for potential social human-machine interactions. Humans are prone to treat tools or toys as substitutes for social agents in everyday interactions. A subjective feeling is not yet sufficient to justify attributing socio-cognitive abilities. For a justified attribution, I argue, *it must be shown that artificial agents contribute to the interactions utilizing socio-cognitive abilities*. By fulfilling minimal necessary conditions, artificial agents can prove as active participants contributing to social interactions.

If both parties actively shape certain human-machine interactions, the observed interaction can no longer be described as mere tool use. This leads to the question of specifying conditions for artificial agents' socio-cognitive abilities and thereby expanding the current conceptual framework of sociality. Overall, we are not only dealing with a conceptual question concerning the conditions that artificial agents must fulfill but also with an epistemological question, namely when we are entitled to conclude that these conditions are met.

Of course, one could still decide to expand the conceptual framework of tool use (e.g., by introducing social tools) instead of revising the conceptual framework regarding social agency. I suspect, however, that the notion of tool use would turn out to be inadequate to capture precisely social aspects of

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

such interactions, such as the *reciprocal exchange of social information*. Assuming that certain interactions with artificial agents are somehow located in a *terra incognita*, where the terminology from philosophy of mind cannot conceptualize them, I will argue below that these phenomena have enough similarities[2] with social interactions between humans. Thus it seems reasonable to opt for re-examining the conceptual framework of social agency. My proposal for the conceptual clarification of the minimal necessary conditions in conjunction with an investigation of epistemological questions aims to provide arguments for claiming that artificial agents can qualify as active participants, respectively, as social agents, in social interactions under certain circumstances.

3 Overcoming restrictive conceptions of sociality

Nevertheless, extending the conception of sociality is a challenging endeavor because it contradicts both our common sense and current philosophical ideas of sociality we can find in philosophy of mind. So far, many conceptions of sociality are limited to living beings. Although the demarcation between the living and the non-living is not always that obvious, up to now, it seems to be widespread that being alive is taken as a necessary but not sufficient condition for social agency. However, assuming that certain interactions with artificial agents should be considered social interactions and not tool use, it seems reasonable to rethink the conditions for social agency.

Turning to philosophy of mind, there are typically thought to be two key elements of social interaction partners: the capacity for genuine agency and certain social abilities. However, on very demanding conceptions, such as those elaborated by Davidson (1980; 2001), neither artificial systems, nor certain disabled persons, nor infants, nor non-human animals fulfill either of the requirements. According to Davidson, the capacity for agency, thought, language, and interpretation are interrelated, and only linguistically sophisticated creatures can be genuine agents, and hence genuine social agents. According to this view, socio-cognitive abilities are characterized as if they were only present in sophisticated adult humans. At this point, it may be debatable how often even human adults actually encounter such ideal cases in everyday life. In addition, one can question the consistently anthropocentric

2 Such similarities are not only grounded in our subjective experiences but also in abilities we are justified ascribing to the artificial agent.

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

character of those concepts. In fact, research indicates that this demanding conception is too demanding. There are multiple realizations of socio-cognitive abilities in various types of agents, such as infants and non-human animals (cf. Brownell 2011; Heyes 2014, 2015; Pacherie 2013; Perler 2005; Premack/Woodruff 1978; Warneken et al. 2006). And different capacities come online at different stages of development (Perner 1991; Tomasello 2008). This evidence supports the common-sense notion that infants are social beings with whom one can interact socially (Vesper et al. 2010), and it might likewise become part of our common sense to consider certain artificial systems as social interaction partners.

What makes me optimistic regarding arguing for an extension of the restrictive conceptual framework in the philosophy of mind is the fact that attitudes toward the status of social agents have proven mutable throughout human history. Here one can point out that, at least as far as the status of women, children, other ethnic groups, and some non-human animals is concerned, an extension of the class of social agents has already arrived in our common sense. This shows that formerly excluded subjects can be considered proper social agents due to social changes.

If we consider the status of slaves, we can even describe a case in which living beings are deprived of their status as social agents and are instead considered as tools without rights[3]. That this should be changed is beyond question. But it vividly illustrates that even the minimal requirement of being alive is not sufficient to maintain social agent status. This is also illustrated by the widespread assumption that one cannot have social interactions with all kinds of non-human animals, e.g., interactions with mosquitoes.

Although it may sound radical to a Western audience, I argue that even non-living beings can be considered social agents. In this context, it is worth mentioning that outside of Western cultural conceptions, e.g., in Shintoism and Animism, there are already conceptions that characterize objects as animate that are considered inanimate from a Western perspective (see Jensen/Blok 2013).

3 For example, Aristotle described slaves as "animate tools" in "Politics" (1, 1253b15-55b40).

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - CC BY-SA

3.1 Towards a broader conception of sociality

Starting from the assertion that the current terminology in philosophy of mind cannot adequately describe certain human-machine interactions, one has to expand the conceptual framework to capture multiple realizations of social agency.

So far, most objections concerning the classification of artificial agents as social agents are based on arguments claiming that non-living entities lack essential abilities which are necessary for social agency (Nida-Rümelin 2022). In this context, for example, the lack of full-fledged intentionality, free will, and emotional states is cited as a reason for excluding artificial agents (Davidson 1980). In the debates about moral agency, the lack of phenomenal consciousness and, in particular, the lack of the ability to suffer is often taken as a reason why artificial agents cannot be moral objects, respectively moral subjects.

However, following the strategy of above mentioned minimal approaches, one can question the necessity of some conditions by demonstrating that multiple realizations of socio-cognitive abilities are conceivable and part of everyday life. For instance, one feature of social interactions that can be elaborated on is that the individuals involved are able to attribute mental states to each other. This ability of mindreading enables them to anticipate, to some extent, the behavior of their interaction partners; if they were not able to do so, social interactions would become immensely difficult, if not impossible. Similar to Davidson's notion of action, one finds very presuppositional notions of mindreading that exclude many types of agents from the outset (Fodor 1992). In response to the restrictive attribution of full-fledged mindreading, Butterfill and Apperly (2013) developed the notion of *minimal mindreading* by questioning the necessity of overly demanding cognitive resources. Thus, they can characterize automated mindreading in adults as well as abilities of children and other animals (cf. Strasser 2012). Since *minimal mindreading* does not require conscious reasoning, this notion is also suitable, as I will argue, to characterize potential abilities of artificial agents.

3.2 Joint action

Since there are manifold social interactions that I cannot consider in-depth, I decided to focus here on a specific subclass, namely joint actions. One might object that acting jointly is one of the most presuppositional form of a social

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

interaction and that it would be more obvious to start with not-so-demanding forms of social interactions. However, I think that if one can show that it is arguable that artificial systems even can qualify as possible partners in joint actions, then the claim that artificial agents can qualify as social agents is on safer ground.

Joint actions are social interactions in which (at least) two agents cooperate and do things together to reach a common goal. According to Bratman, having a shared intention is taken as the essential condition of joint actions among human adults (Bratman 2014: 152). However, the fulfillment of the conditions for having shared intention proves to be demanding. Just having an intention is not enough. In addition, one has to be able to entertain a specific belief state that enables a relation of interdependence and mutual responsiveness between one's own intentions and the others. In short, acting jointly, you need shared goals. All this, it is assumed, requires the ability to have common knowledge, mastery of mental concepts, and sophisticated mentalization skills. Following Bratman, disabled persons, children, and non-human animals are excluded because they lack sophisticated mental concepts and capacities for explicit commitments. However, this conflicts with both our common sense and empirical data (cf. Brownell 2011). Children are understood as socially interacting beings even though they do not fulfill the demanding conditions of a standard notion of joint action (cf. Vesper et al. 2010). For example, playing hide and seek with children is experienced as a proper joint action. Furthermore, research indicates that not only children but also non-human animals successfully engage in joint actions without fulfilling the demanding conditions of Bratman's notion (cf. Warneken et al. 2006). Asking whether artificial agents might be able to participate in a joint action is just one step further.

The strategy of minimal approaches to capture a broader range of socio-cognitive abilities by proposing minimal versions of established notions is, in my view, a promising starting point to overcome restrictive conceptions of sociality. By assuming multiple realizations, I can offer an extension of the restrictive notions in play and present a way how one can also capture socio-cognitive phenomena with respect to artificial agents. If one can establish the idea of multiple realizations, one can question the absoluteness of the demanding conditions put forward. In the best case, I can thus show on what basis a further discussion of these nevertheless strikingly restrictive theoretical approaches continues to make sense. Instead of assuming full-fledged joint actions in which all participants satisfy the same demanding

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

conditions, one can then show that not all the conditions we require in the human case, such as having emotions, turn out to be necessary for artificial agents as well. On the basis of multiple realizations, one can work out a set of minimal necessary conditions that satisfy the requirements we impose on artificial agents in social interactions.

4 A minimal notion of joint action

To develop a minimal notion of joint action, I start with a rough working definition: Every agent acting jointly must be able to act and coordinate. The supposed necessity of particular involved requirements regarding coordination will be investigated step by step.

4.1 Asymmetric actions

Joint actions involving children or non-human animals already hint at the possibility of multiple realizations in which those conditions can be fulfilled. Due to multiple realizations, the distribution of abilities can vary among the participants. This is what I call asymmetric joint actions. For example, mother-child interactions illustrate that the distribution between participants is not necessarily equal.

Assuming that there are also multiple realizations regarding artificial agents, it is only one step further to suppose that besides joint actions with mixed groups consisting of human adults and children, there may as well be joint actions with humans and artificial agents. Describing human-machine interactions as asymmetric joint actions, there is no need to suppose the very same sets of conditions for artificial agents. Returning to my rough definition of joint action, saying that every agent acting jointly must be able to act and coordinate, the conditions of a *minimal joint action* will presuppose multiple realizations of acting and coordinating, respectively *minimal agency* and *minimal coordination*. Focusing on asymmetric joint actions, the developed notion clarifies which conditions should be imposed on the different participants of such joint actions. A task for future research would be to investigate whether joint actions can also occur between two or more artificial agents, i.e., whether there can be social interactions in which no humans are involved.

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

4.2 Ability to act

Without a doubt, if you are not able to act, you cannot act jointly. According to philosophy of mind, one of the important features of actions is that they are "intentional under some description" (Davidson 1980: essay 3). However, due to a strong sense of intentionality, this implies various further abilities, such as being equipped with consciousness, generating goals, and making free choices. Full-fledged agency, as, for instance, described by Davidson, includes highly demanding conditions. Besides intentionality, consciousness, the ability to generate goals, and the ability to make free choices, it is also required that acting agents have propositional attitudes and a mastery of language. According to such demanding conditions, non-living beings, non-human animals, people with disabilities, and children cannot act. All that they are capable of is producing more or less complex behavior. Criticism concerning the exclusion of children and non-human animals has already been raised. What has to be shown now is that this critique can also concern the exclusion of non-living artificial agents.

Assuming that the abilities of artificial agents in certain human-machine interactions are neither adequately described by mere behavior nor by full-fledged agency, I argue that one should make a finer-grained differentiation concerning the classes of events, such as behavior and action. With a notion of a *minimal action*, one can capture phenomena in-between mere behavior and full-fledged actions (cf. Strasser 2006). On the basis of a minimal notion, one can then argue in a similar way as Wallach and Allen (2009) do with respect to moral agency, that agency should be understood as a gradual concept. Minimal and full-fledged agency could then characterize the extreme cases of a continuum. In line with the rationale of other minimal approaches, the notion of a *minimal action* can question the necessity regarding some conditions that exclude artificial agents from the outset. Assuming that agency can be reached by interpreting proposed conditions in a weaker sense, the necessity of a full-fledged realization of all requirements is questioned. For example, one can question whether *minimal agency* necessarily requires that the generation of goals occurs in the acting entity. Alternatively, a goal can be generated in another system and then transferred to the minimal acting system. If the minimal acting system can recognize and represent goals as goals, it can still act goal-directed. Likewise, being conscious can turn out to be not a necessary condition. This is not to deny that consciousness plays an important role in human cases. But given that there are multiple realizations, consciousness

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

could be a specific property of living beings (a biological constraint) that lacks necessity with respect to artificial agents.

Roughly speaking, *minimal agency* requires that artificial agents are able to perceive, represent, and process the relevant information (including goals, context, etc.) and must have effectors to perform an action. Of course, the development of such a notion requires more specifications, clarifying the extent to which perception, representation, and processing include abilities to adapt and learn (cf. Strasser 2006, 2015). For the sake of argument, let us assume that one can presuppose minimal agency with respect to artificial agents. It may be important to clarify that this minimal notion requires conditions that the minimal actor must actually fulfill, and should not be confused with attribution practices of other actors, such as those described by Dennett's (1987) intentional stance.

Now, when describing asymmetric joint actions, one can specify two different realizations by referring to the minimal and full-fledged notions of agency. Artificial agents can realize the ability to act in a more minimal way that does not require conscious, mental, or emotional states. At the same time, the more demanding conditions of a full-fledged agency can describe the agency of the human counterpart.

4.3 Coordination

Just being able to act is not sufficient to qualify as a participant in a joint action. In addition, the ability to coordinate is an essential prerequisite for joint actions. This ability plays a crucial role for the social dimension of joint actions. With reference to Bratman's notion of a joint action that requires shared intentions, one can argue that the functional role of coordination is to enable shared intentions. Only if agents work together in an organized way, we can talk of joint actions. Investigating minimal necessary conditions artificial agents have to fulfill to coordinate with human counterparts, I elaborate on three important aspects: reciprocal exchange of social information, mindreading, and commitment.

4.3.1 Reciprocal exchanges of social information

Explorations of human social cognition highlight the importance of social signals (cf. Frith/Frith 2007). Working together in an organized way requires reciprocal exchanges of social information. In human-human interactions, we observe an exchange of a wealth of information transferred by language or

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

other expressive behaviors. In addition to verbal agreements, also social cues such as gestures and facial expressions are exchanged. Deficits in interpreting non-verbal behavior can lead to deficits in social interactions (cf. Mundy et al. 1986; Bogart/Tickle-Degnen 2015). Since humans apply social cues in joint actions, it is a necessary requirement for artificial agents to interpret and send social cues back to their human counterparts. Regarding tool use, there is no need for a reciprocal exchange of social information. In contrast, processing social cues seems to be a distinguishing feature of social interactions. Consequently, the first requirement for coordination with human counterparts concerns the ability to handle social cues. I aim to show that it is not necessary that artificial agents actually have emotional and mental states. From the perspective of establishing minimally necessary criteria, I argue that it is sufficient for artificial agents to have the ability to express and interpret social signals. Analogous to the ability to act, it can be argued here that there are multiple realizations of how a condition can be satisfied. For example, instead of requiring emotional or mental states, one could implement functions that are realized by emotional or mental states in the human case. This is in line with Wallach and Allen (2009), who speak of functional equivalence in this context. The general ability to process and interpret social information constitutes social competence, which seems to be an essential prerequisite for any kind of social interaction.[4]

4.3.2 Mindreading

Another relevant aspect of coordination in joint actions consists in the fact that participants normally are able to anticipate to some extent what the other agent will do next. In the humanities and natural sciences, one aspect of such anticipation abilities is discussed under the label *mindreading* or *Theory of Mind* (cf. Fodor 1992; Fletcher/Carruthers 2013; Gopnik 2003, Nichols/ Stich 2003). Once again, there are notions that are tailored to human adults only. However, research indicates that mindreading abilities are present in children and non-human animals. This motivated Butterfill and Apperly (2013) to develop the notion of *minimal mindreading*, which can account for a broader

4 At this point, one might get the impression that the development of minimal notions does take place in the spirit of Dennett's intentional stance (1987). However, I think there is a difference between requiring conditions to be met by an entity in order to be justified in attributing an ability to that entity and adopting an intentional stance because it is practicable.

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

range of mindreading. Questioning the necessity of overly demanding cognitive resources, such as representing a full range of complex mental states and a mastery of language, they elaborated minimal necessary conditions that can explain success in mindreading tasks. For example, the full range of representations of complex mental states is replaced by representations of less complex mental states, specified as encounterings and registrations, and it is argued that they are sufficient to anticipate the behavior of other agents in an efficient, automatic, fast, and robust manner (Butterfill/Apperly 2013: 18). Most significantly, with regard to artificial agents, *minimal mindreading* does not require conscious reasoning. Thus, the notion of *minimal mindreading* can account for mindreading abilities, such as automatic mindreading in human adults and mindreading abilities in children and non-human animals. Therefore, the second requirement for coordination with human counterparts concerns the ability to display *minimal mindreading* abilities.

4.3.3 Commitment

Besides processing social information and anticipating the behavior of the counterpart, it is crucial for the success of human-human joint actions that both agents can rely on the contribution of the other agent. This means both parties are committed to sticking to the joint action in the ideal case. Another important function of commitments is that they make agents' behavior easier to predict (Michael/Pacherie 2015: 100). It is not surprising that some notions of a (strict) commitment are tailored to human adults only (cf. Shpall 2014). Strict commitments are characterized as a triadic relation among two agents and an action. Involved agents mutually have certain expectations and motivations concerning this action.

To explore in what sense commitment is important for human-machine interactions, the minimal approach by Michael et al. (2016) offers a good starting point. Michael and colleagues argue that components of a strict commitment can be dissociated. They suggest that a single occurrence of one component can already be treated as a sufficient condition for a *minimal sense of commitment*.

Following the idea that components of a commitment can be disassociated, one can describe a *minimal sense of commitment* on the human side with respect to human-machine interactions. Regardless of whether the artificial agent is committed, a human counterpart can have an expectation of what the artificial agent should do and thereby assume that the artificial agent is com-

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

mitted. Alternatively, the human can be motivated to contribute because she implicitly assumes that the artificial agent is expecting this. By disassociating expectations from corresponding motivations, the *minimal sense of commitment* does not rely on the corresponding counterpart's abilities.

Focusing on the side of the artificial agents, one might object that we are not justified to ascribe motivations or expectations to artificial agents. Therefore, analyzing the side of the artificial agents with respect to a *minimal sense of commitment* is challenging. Avoiding the requirement of mental and emotional states, such as expectation and motivation, I suggest allowing functional corresponding states of expecting and feeling motivated (cp. functional equivalence Wallach & Allen 2009: 68). Thereby, I follow the strategy of questioning whether certain biological constraints are necessary. Formulating conditions of how a *minimal sense of commitment* can be realized in human-machine interactions, the third requirement for artificial agents demands an ability to interpret signs of being committed regarding the human counterpart as well as the ability to react by signaling expectations and motivations in order to contribute to the joint action.

In sum, to qualify as a proper participant in a joint action with a human counterpart, artificial agents have to fulfill conditions ensuring that we can ascribe *minimal agency* and an active contribution to the coordination of this joint action. Regarding coordination, the artificial agent has to be able to join a reciprocal exchange of social information. Furthermore, artificial agents should display *minimal mindreading abilities* as well as the ability to exhibit and elicit a *minimal sense of commitment*.

5 Pieces of the puzzle found in AI

So far, the development of a notion of *minimal joint actions* has remained on a theoretical, conceptual level. Approaching epistemological questions as to whether we are justified in ascribing artificial agents the required abilities, one has to evaluate actual abilities of artificial agents. The following examples show that, though distributed over distinct systems, proposed conditions can be fulfilled in principle. Consequently, it seems conceivable that even a single artificial agent may fulfill all conditions to be a participant in a *minimal joint action* in the near future. Nevertheless, one has to admit that the claim that artificial agents could be proper participants in social interactions holds only for limited situations at this point.

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - CC BY-SA

Assuming that artificial agents can qualify for *minimal agency* (cf. Strasser 2006), the elaborated conditions for coordination, namely exchanging social information, *minimal mindreading*, and a *minimal sense of commitment*, stand in the center of this brief investigation. Since every form of communication is a joint action, communication can serve as a prototypical case to investigate to what extent artificial agents are able to contribute to a joint action. Analyzing face-to-face communication as a joint action, mutual understanding can be described as a shared goal of communicating agents. Both agents contribute via language and expressive behaviors (such as facial expressions, gestures, body postures, or prosody) to an exchange of information in order to reach mutual understanding. Particularly, they can make their minds visible by expressing social cues.

5.1 Reciprocal exchange of social information

According to the minimal necessary conditions of the ability to act jointly, both participants must be able to exchange social information. Describing a reciprocal exchange of social information in human-machine interactions, one can, for example, imagine that a human is saying 'Hi' to her artificial counterpart and sends a social cue, such as a smile, and then the artificial agent may smile back saying 'Hi, nice to see you.' To this end, the artificial agent has to be able to detect and process both the linguistic expression and the expressive behavior of the human.

For example, deep learning networks for emotion recognition can be used to recognize emotional expressions (cf. Mossbridge/Monroe 2018). Furthermore, showing that artificial agents are able to process such information, they have to be able to respond with appropriate answers, which are interpretable by humans. Otherwise, artificial agents cannot participate in a reciprocal exchange of social information. The ability of artificial agents to express social cues might not yet be that differentiated. However, the human tendency to anthropomorphize will help regarding interpreting artificial agents' gestures and emotional expressions. A nice example of how social information is exchanged is the artificial agent *Max* – a virtual human developed by Ipke Wachsmuth and his team (2008) – who can give rise to secondary emotions such as frustration and relief. This is realized by a belief-desire-intention-based cognitive module in which an emotion dynamics simulation system is integrated (cf. Becker/Wachsmuth 2006; an overview of the research of emotional expression can be found in Petta et al. 2011). Furthermore, there are also

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

artificial agents that are able to interpret social cues such as gestures (Kang et al. 2012). *Artificial Retrieval of Information Assistants* (ARIAs) that are able to handle multimodal social interactions (Baur et al. 2015) demonstrate how linguistic and expressive behavior can be brought together. They can maintain a conversation with a human agent and, indeed, react adequately to verbal and nonverbal behavior.

Technically speaking, artificial agents have appropriate detection systems, reasoning mechanisms, and the ability to express social cues to participate in a reciprocal exchange of information.

5.2 Minimal mindreading

With respect to acting jointly with a human counterpart, I claimed that reasoning mechanisms should entail *minimal mindreading* abilities. There is no doubt that artificial agents possess abilities that can be understood as reasoning mechanisms. Concerning mindreading abilities, one can refer to the work of Gray and Breazeal (2014). They presented an artificial agent that is able to model mental states concerning the perspective of a human counterpart. This agent is able to infer from its perception of the physical world to what a human counterpart can see or cannot see. Moreover, it is able to consider that the perspective of the counterpart will direct future actions of the human. To illustrate the capabilities of this artificial agent, we can imagine a situation where this artificial agent can see three entities (two objects and a human), while the human can only see two entities (one object and the artificial agent) because the second entity is behind the human and therefore out of sight. First, the artificial agent constructs a model of its world perspective, then the human position and orientation in this model are used to convert incoming sensor data into data that are relative to the human coordinate system. Thereby, entities that are not visible to the human are filtered out, and the model is transformed into a human-centric format. This model can then be used to anticipate future human behavior. Even though such capabilities are so far only valid in a limited range of situations, this artificial agent is not only capable of modeling mental states with respect to the perspective of a counterpart, but it is also able to use such perspectival aspects as a factor to anticipate human behavior. Therefore, one can describe this ability as a multiple realization of mindreading – as *minimal mindreading*.

Another example demonstrating how artificial agents make use of their detection and reasoning systems in order to anticipate future behavior of a

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

counterpart can be found in the work of Cavallo and colleagues (2016). They show that one can train a classification and regression tree model (CART) in order to read intentions. Of course, this is again limited to a very specific subclass of behavior, namely predicting which intention a specific movement has. To make a long story short, neuroscientific research about the role of implicitly processed information in movement kinematics suggests that humans are very good at predicting future actions. For example, it shows that humans are able to predict at a quite early point of time whether an observed agent is grasping a bottle with the intent to pour water into a glass or to drink water from the bottle (Cavallo et al. 2016; Manera et al. 2011; Sartori et al. 2011). Based on this research, Cavallo and colleagues trained and tested a CART model, which was fed with kinematic information using various sensors. The accuracy of this CART model in predicting the intentions of human counterparts is already impressively high.

5.3 Minimal sense of commitment

Last but not least, I suggested that *minimal joint actions* require a *minimal sense of commitment*. A critical issue concerns the question of whether both types of agents – humans and machines – have to display a minimal sense of commitment. Considering psychological factors, such as the human tendency to anthropomorphize, it seems well possible that human agents establish a *minimal sense of commitment* towards artificial agents. Humans can be motivated to stick to a joint action because they assume that their counterpart is expecting this. Moreover, they are also able to expect (project) that the artificial counterpart is motivated by a sense of commitment.

With respect to a *minimal sense of commitment* at the side of the artificial agents, things get a little bit more complicated. Presuming that artificial agents can learn to interpret signs of their human counterparts' commitment, they could react adequately by also signaling expectations and motivations to contribute to the joint action without actually having expectations or motivations. To this end, one can refer to research projects that implement expressive communication in order to enhance trust (Hamacher 2016).

What remains questionable is whether these functional equivalent forms of a minimal sense of commitment can guarantee reciprocity all the way through. However, if we take the claim that a *minimal sense of commitment* can be disassociated seriously, it might be sufficient if the human counterpart takes over the commitment task.

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - CC BY-SA

6 Conclusion

Leaving aside the psychological fact that many human-machine interactions appear as they were social interactions, I investigated the minimal conditions which should be fulfilled by artificial systems to qualify as social agents using joint actions as a demanding example of a social interaction. According to my view, the question as to whether artificial agents might be able to enter the realm of social cognition by qualifying as social agents that are able to act jointly with humans concerns both conceptual and epistemological issues. Given that reciprocal exchanges of social cues mark a distinguishing feature of social interactions, interactions in which artificial agents *contribute* to such an exchange should not be reduced to mere tool use. Consequently, the conceptual framework of tool use is not sufficient to account for such human-machine interactions. However, especially in the debates about joint actions in the philosophy of mind, there are no established concepts to capture socio-cognitive phenomena of artificial agents. This is why I diagnosed a need to review and expand the conceptual framework.

To this end, I delivered a sketch of how to develop a notion of *minimal joint action* which is applicable to artificial agents. Claiming that agency and coordination are essential for the ability to act jointly, this proposal questions whether socio-cognitive abilities are necessarily based on biological constraints. Instead of full-fledged agency with full-blown autonomy and consciousness, it is argued that *minimal agency* is sufficient. With respect to coordination, it is argued that if artificial agents are able to process social information, they can succeed in fulfilling conditions, such as reciprocal exchange of social information, mindreading, and maybe even commitment, which all contribute to the ability of coordination. By outlining minimal necessary conditions artificial agents have to fulfill in order to qualify as a new type of social participants in joint actions, I argue that artificial agents can, to a limited extent, establish reciprocal exchanges of social information and thereby qualify as social agents. Pointing to some research achievements in AI, I demonstrated that the potential fulfillment of proposed conditions is not out of reach. Where previous revolutions have dramatically changed our environments, this one has the potential to change our understanding of sociality substantially.

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - CC BY-SA

References

Aristotle (1905): Aristotle's Politics. Oxford: Clarendon Press; http://data.perseus.org/citations/urn:cts:greekLit:tlg0086.tlg035.perseus-eng1:1.

Baur, Tobias/Mehlmann, G./Damian, I./Gebhard, P./Lingenfelser, F./Wagner, J./Lugrin, B./André, E. (2015): "Context-Aware Automated Analysis and Annotation of Social Human-Agent Interactions." In: ACM Transactions on Interactive Intelligent Systems 5/2, pp. 11:1-11:33.

Becker, Christian/Wachsmuth, Ipke (2006): "Modeling primary and secondary emotions for a believable communication agent." In: D. Reichardt/P. Levi/J.J.C. Meyer (eds.), Proceedings of the 1st Workshop on Emotion and Computing, pp. 31-34.

Bogart, Kathleen R./Tickle-Degnen, Linda (2015): "Looking beyond the face: A training to improve perceivers' impressions of people with facial paralysis." In: Patient Education and Counseling 98, pp. 251-256.

Bratman, Michael (2014): Shared agency: A planning theory of acting together, Oxford: Oxford University Press.

Brownell, Celia A. (2011): "Early Developments in Joint Action." In: Review of philosophy and psychology 2/2, pp. 193-211.

Bryson, Joanna J. (2010): "Robots should be slaves." In: Yorick Wilks (ed.), Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues, pp. 63-74.

Butterfill, Steven/Apperly, Ian (2013): "How to construct a minimal theory of mind." In: Mind and Language 28/5, pp. 606-637.

Carpenter, Julie (2016): Culture and human-robot interaction in militarized spaces: a war story, London: Routledge.

Cavallo, Andrea/Koul, Atesh/Ansuini, Caterina et al. (2016): "Decoding intentions from movement kinematics." In: Scientific Reports 6, 37036.

Damiano, Luisa/Dumouchel Paul (2018): "Anthropomorphism in Human-Robot Co-evolution." In: Frontiers in Psychology 9, 468.

Darling, Kate (2016): "Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior toward robotic objects." In: R. Calo/A. M. Froomkin/I. Kerr (eds.), In Robot law, Northampton, MA: Edward Elgar, pp. 213-231.

Darwall, Stephen (2006): "The Value of Autonomy and Autonomy of the Will" In: Ethics 116, pp. 263-284.

Davidson, Donald (1980): Essays on actions and events, Oxford: Oxford University Press.

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - CC BY-SA

Davidson, Donald (2021): Subjective, Intersubjective, Objective. Oxford: Oxford University Press.

Dennett, Daniel (1987): The intentional stance, Cambridge, MA: MIT Press.

Fletcher, Logan/Carruthers, Peter (2013): "Behavior-Reading versus Mentalizing in Animals." In: Janet Metcalfe/Herbert Terrace (eds.), Agency and Joint Attention, Oxford: Oxford University Press, pp. 82-99.

Floridi, Luciano/Sanders, J.W. (2004): "On the Morality of Artificial Agents." In: Minds and Machines 14, pp. 349-379.

Fodor, Jerry (1992): "A Theory of the Child's Theory of Mind." In: Cognition 44/3, pp. 283-296.

Frith, Chris/Frith, Uta (2007): "Social cognition in humans." In: Current Biology 17/16, pp. R724-R732.

Gopnik, Alison (2003): "The Theory Theory as an Alternative to the Innateness Hypothesis." In: Louise M. Antony (ed.), Chomsky and His Critics, Malden, MA: Blackwell, pp. 238-254.

Gray, Jesse/Breazeal, Cynthia (2014): "Manipulating Mental States Through Physical Action – A Self-as-Simulator Approach to Choosing Physical Actions Based on Mental State Outcomes." In: International Journal of Social Robotics 6/3, pp. 315-327.

Hakli, Raul/Seibt, Johanna (2017): Sociality and Normativity for Robots: An Introduction, Cham: Springer, pp. 1-10.

Hamacher, Adriana/Bianchi-Berthouze Nadia/Pipe Anthony G./Eder Kerstin (2016): "Believing in BERT: Using expressive communication to enhance trust and counteract operational error in physical Human-Robot Interaction." In: 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp. 493-500.

Heyes, Cecilia (2014): "False belief in infancy: a fresh look." In: Developmental Science, 17/5, pp. 647-659.

Heyes, Cecilia (2015): "Animal mindreading: what's the problem?" In: Psychonomic Bulletin/Review 22/2, pp. 313-327.

Hortensius, Ruud/Cross, Emily S. (2018): "From automata to animate beings: the scope and limits of attributing socialness to artificial agents." In: Annals of the New York Academy of Sciences, pp. 1-18.

Jensen CB/Blok A. (2016): "Techno-animism in Japan: Shinto Cosmograms, Actor-network Theory, and the Enabling Powers of Non-human Agencies." In: Theory, Culture & Society 30/2, pp. 84-115.

Johnson, Deborah G. (2006): "Computer systems: Moral entities but not moral agents." In: Ethics and Information Technology 8/4, pp. 195-204.

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - CC BY-SA

Kang, S./Gratch, J./Sidner, C./Artstein, R./Huang, L./Morency, L.P. (2012): "Towards building a Virtual Counselor: Modeling Nonverbal Behavior during Intimate Self-Disclosure." In: Eleventh International Conference on Autonomous Agents and Multiagent Systems 1, pp. 63-70.

Leveringhaus, Alex (2016): Ethics and Autonomous Weapons, London: Palgrave Macmillan UK.

List, C. (2021): "Group Agency and Artificial Intelligence." In: Philosophy & Technology, 34, pp. 1213-1242.

Loh, Janina (2019): Roboterethik: Eine Einführung, Frankfurt am Main: Suhrkamp.

Manera, Valeria/Becchio, Cristina/Cavallo, Andrea/Sartori, Luisa/Castiello, Umberto (2011): "Cooperation or competition? Discriminating between social intentions by observing prehensile movements." In: Experimental Brain Research 211, pp. 547-556.

Michael, John/Pacherie, Elisabeth (2015): "On Commitments and Other Uncertainty Reduction Tools in Joint Action." In: Journal of Social Ontology 1/1, pp. 89-120.

Michael, John/Sebanz, Nathalie/Knoblich, Günther (2016): "The Sense of Commitment: A Minimal Approach." In: Frontiers in Psychology 6, 1968.

Misselhorn, Catrin (2018): Grundfragen der Maschinenethik, Stuttgart: Reclam.

Mossbridge, Julia/Monroe, Edward (2018): "Team Hanson-Lia-SingularityNet: Deep-learning Assessment of Emotional Dynamics Predicts Self-Transcendent Feelings During Constrained Brief Interactions with Emotionally Responsive AI Embedded in Android Technology." In: Unpublished XPrize Submission (downloadable at: https://www.academia.edu/38677142).

Mundy, Peter/Sigman, Marian/Ungerer, Judy/Sherman, Tracy (1986): "Defining the social deficits of autism: the contribution of non-verbal communication measures." In: Journal of Child Psychology and Psychiatry 27/5, pp. 657-669.

Nichols, Shaun/Stich, Stephen (2003): Mindreading. An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds, Oxford: Oxford University Press.

Nida-Rümelin Julian (2022): "Digital Humanism and the Limits of Artificial Intelligence." In: Werthner H./Prem E./Lee E.A./Ghezzi C. (eds.), Perspectives on Digital Humanism, Cham: Springer, pp. 71-75.

Pacherie, Elisabeth (2013): "Intentional joint agency: shared intention lite." In: Synthese 190/10, pp. 1817-1839.

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Perler, Dominik/Wild, Markus (2005): Der Geist der Tiere. Philosophische Texte zu einer aktuellen Diskussion, Frankfurt am Main: Suhrkamp.

Perner, Josef (1991): Understanding the Representational Mind, Cambridge, MA: MIT Press.

Petta, Paolo/Pelachaud, Cathrin/Cowie, Roddy (2011): Emotion-Oriented Systems: The Humaine Handbook, Heidelberg/Dordrecht/London/New York: Springer.

Premack, David/Woodruff, Guy (1978): "Does the chimpanzee have a theory of mind?" In: Behavioral Brain Sciences 1, pp. 515-526.

Sartori, Lucy/Becchio, Cristina/Castiello, Umberto (2011): "Cues to intention: The role of movement information." In: Cognition 119, pp. 242-252.

Shpall, Sam (2014): "Moral and rational commitment." In: Philosophy and Phenomenological Research 88/1, pp. 146-172.

Strasser, Anna (2006): Kognition künstlicher Systeme, Frankfurt: Ontos.

Strasser, Anna (2012): "How Minimal Can Self-Consciousness Be?" In: Grazer Philosophische Studien 84/1, pp. 39-62.

Strasser, Anna (2015): "Can artificial systems be part of a collective action?" In: Catrin Misselhorn (ed.), Collective Agency and Cooperation in Natural and Artificial Systems. Explanation, Implementation and Simulation. Series: Philosophical Studies Series, Vol. 122. Berlin: Springer, pp. 205-218.

Tomasello, M (2008): Origins of Human Communication, Cambridge, MA: MIT Press.

Vesper, Cordelia/Butterfill, Steven/Sebanz, Nathalie/Knoblich, Günther (2010): "A minimal architecture for joint action." In: Neural Networks 23/8/9, pp. 998-1003.

Wachsmuth, Ipke (2008): "'I, Max' – Communicating with an artificial agent." In: Ipke Wachsmuth/Günther Knoblich (eds.), Modeling Communication with Robots and Virtual Humans. Lecture Notes in Computer Science, Vol. 4930, Berlin: Springer, pp. 279-295.

Wallach, Wendell/Allen, Collin (2009). Moral machines: Teaching robots right from wrong, Oxford: Oxford University Press.

Warneken, Felix/Chen, Frances/Tomasello, Michael (2006): "Cooperative activities in young children and chimpanzees." In: Child development 77/3, pp. 640-663.

Wykowska, Agnieszka/Chaminade, Thierry/Cheng, Gordon (2016): "Embodied artificial agents for understanding human social cognition." In: Philosophical transactions of the Royal Society of London. Series B, Biological sciences 371/1693, 20150375.

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

https://doi.org/10.14361/9783839462652-004 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA